

SISTEM UNTUK MENGANALISA FEEDBACK MAHASISWA TERHADAP PROSES PERKULIAHAN DENGAN METODA NAÏVE BAYES CLASSIFFIER

Herianto¹, Gita Fitriana²

¹ Dosen Jurusan Teknik Informatika Fakultas Teknik, Universitas Darma Persada

² Jurusan Teknik Informatika Fakultas Teknik, Universitas Darma Persada

Abstrak

Penelitian ini bertujuan untuk membangun sistem yang dapat melakukan analisa sentimen feedback mahasiswa pada portal di universitas darma persada berupa kepuasan mahasiswa dalam proses perkuliahan. Analisis sentimen adalah proses klasifikasi data yang dalam penelitian ini dibagi menjadi dua bagian yaitu yang bersifat positif dan yang bersifat negatif. Klasifikasi sangat penting dalam menentukan jenis feedback yang diberikan dalam bentuk komentar. Proses klasifikasi diawali dengan membagi koleksi data menjadi data latih dan data uji. Data latih digunakan metode naïve bayes classifier sehingga diperoleh model klasifikasi untuk penentuan kelas pada data uji. Metode naïve bayes classifier adalah metode yang berdasarkan probabilitas dan teorema bayesian. Metode ini digunakan untuk mengklasifikasikan hasil data freedback mahasiswa yang dituliskan pada kuesioner mahasiswa terhadap dosen sehingga menghasilkan klasifikasi secara otomatis yang diinginkan. Hasil penelitian ini digunakan untuk menentukan klasifikasi feedback dari data kuesioner mahasiswa sehingga data mudah terbaca.

Keyword : Analisis feedback, Naïve Bayes Classifier, Kuesioner Mahasiswa.

1. PENDAHULUAN

1.1 Latar Belakang

Kualitas pengajaran dan standarisasi akademik perlu dilakukan evaluasi dan ditingkatkan setiap saat untuk menghasilkan mahasiswa yang berkualitas. Pengajaran yang baik adalah sesuatu yang sulit dan tidak mudah, tergantung kepada para pengajar yang berkaitan langsung dengan proses perkuliahan, maka evaluasi kepada para pengajar sangat diperlukan untuk mengidentifikasi kontribusi yang sudah dilakukan dalam pencapaian tujuan pengajaran.

Universitas Darma Persada sebagai lembaga pendidikan tinggi senantiasa ingin meningkatkan layanan dalam manajemen pembelajaran. Untuk maksud tersebut pada setiap akhir semester, akademik mengadakan evaluasi layanan pembelajaran menggunakan instrumen kuesioner dengan butir-butir jawaban yang telah disediakan. Selama ini ada data kuesioner yang tidak dapat dimanfaatkan dan dianalisis yaitu data saran mahasiswa. Untuk menganalisa data-data saran/opini mahasiswa, sehingga saran-saran dan opini tersebut dapat dimanfaatkan untuk mendukung hasil evaluasi menggunakan skor kuesioner.

Berdasarkan penjabaran di atas peneliti tertarik membuat aplikasi yang dapat mengklasifikasikan feedback komentar mahasiswa di SIAKAD Unsada apakah

termasuk opini yang positif atau negatif menggunakan algoritma Naïve Bayes Classifier.

1.2. Rumusan Masalah

Rumusan masalah dalam penelitian ini adalah :

Bagaimana feedback komentar mahasiswa dalam kuesioner dapat dianalisa dengan Naïve Bayes Classifier dan sejauhmana tingkat akurasi yang dihasilkan ?

1.3. Batasan Masalah

Batasan masalah pada penelitian ini :

- a. Feedback komentar hanya dapat diproses bila diisi dengan menggunakan Bahasa Indonesia yang dan sesuai EYD, tidak berupa singkatan, tidak berupa bahasa gaul dan tidak menggunakan bahasa daerah.
- b. Hasil Klasifikasi hanya berupa perbandingan sentimen positif dan negatif.

1.4. Tujuan Penelitian

Tujuan Penelitian ini adalah :

- a. Mendapatkan informasi penting dalam setiap klasifikasi
- b. Menguji akurasi metode Naïve Bayes Classifier dalam analisa sentimen komentar.

1.5. Manfaat Penelitian

Manfaat yang dapat dipetik dalam penelitian ini antara lain adalah sebagai berikut :

1. Memperoleh gambaran terhadap data yang berupa teks dari bentuk kualifikasi ke bentuk kuantisasi.
2. Hasil penelitian ini diharapkan dapat membantu memberikan informasi analisa sentimen untuk mengetahui respon dari mahasiswa yang tidak terstruktur dengan proses teks mining, sehingga dapat dijadikan sebagai acuan dalam upaya menjaga kualitas dan memperbaiki kekurangan serta evaluasi ke arah yang lebih baik.

1.6. Metode Pengumpulan Data

Penelitian ini memperoleh data dengan cara :

- Observasi : Metode pengumpulan data dengan melakukan pengamatan langsung terhadap objek yang akan kita teliti. Metode ini dapat memperoleh gambaran yang menyeluruh secara relevan.
- Wawancara : Metode pengumpulan data dan fakta penting yang dilakukan dengan cara memberikan pertanyaan langsung kepada pihak – pihak terkait dengan data yang diperlukan oleh aplikasi yang akan dibangun.
- Studi Pustaka : Metode ini dilakukan dengan cara membaca dan memperoleh data yang ada dengan mempelajari buku – buku referensi yang terdapat di perpustakaan dan media lain yang dapat menunjang hasil laporan yang berkaitan dengan bahasa pemrograman yang digunakan.

2. LANDASAN TEORI

Metoda Naïve Bayes Classifier (NBC)

Metode NBC menempuh dua tahap dalam proses klasifikasi teks, yaitu tahap pelatihan dan tahap klasifikasi. Probabilitas adalah kemungkinan terjadinya suatu peristiwa antara 0 s/d 1. Menurut Hadna, N. M., Santosa, P.I., & Winarno (Studi

Literatur tentang perbandingan metode untuk proses analisis sentimen di Twitter, (2016) mengemukakan bahwa Metode *Naive Bayes Classifier* (NBC), Bergantungnya pada model *probabilitas* nya, naive bayes classifier dapat dilatih untuk melakukan supervised learning dengan sangat efektif. Dalam berbagai macam penerapannya, estimasi parameter untuk model naive bayes menggunakan metode maximum likelihood, yang artinya pengguna dapat menggunakan model naive bayes tanpa perlu mempercayai probabilitas bayesian atau tanpa menggunakan metode bayesian.

menghitung prior :

$$P(v_j) = \frac{\text{docs } j}{\text{training}} \quad (1)$$

Dimana :

docs j : jumlah dokumen pada kategori j

training : jumlah dokumen yang digunakan dalam proses *training*

$p(v_j)$ dan probabilitas kata a_1 untuk setiap kategorinya $P(a_i|v_j)$ dihitung pada saat training yang dirumuskan sebagai berikut :

$$P(a_i|v_j) = \frac{n_{i+1}}{n + \text{kosa kata}} \quad (2)$$

Dimana :

n_i : jumlah kemunculan kata a_i pada kategori v_j

n : jumlah kosakata yang muncul pada kategori v_j

kosakata : jumlah kata unik pada semua data *training*

Naive bayes classifier menyederhanakan hal ini dengan mengasumsikan bahwa didalam setiap kategori, Dengan kata lain :

$$V_{MAP} = \text{argmax}_{v_j} \in P(v_j) \times \pi_i P(a_i|v_j) \quad (3)$$

Evaluasi untuk metoda datamining untuk proses klasifikasi umumnya dilakukan menggunakan sebuah himpunan data uji, yang tidak digunakan dalam pelatihan klasifikasi tersebut, dengan suatu ukuran tertentu. Terdapat sejumlah ukuran yang dapat digunakan untuk menilai atau mengevaluasi model klasifikasi, diantaranya adalah: *accuracy* atau tingkat pengenalan, *error rate* atau tingkatan.

Jika data positif dan diprediksi positif akan dihitung sebagai *true positive*, tetapi jika data itu diprediksi negatif maka akan dihitung sebagai *false negative*. Jika data negatif dan diprediksi negatif akan dihitung sebagai *true negative*, tetapi jika data tersebut diprediksi positif maka akan dihitung sebagai *false positive* Hasil klasifikasi biner pada suatu dataset dapat direpresentasikan dengan matriks disebut confusion matrix.

$$\text{Akurasi} = \frac{TP+TN}{(TP+TN+FP+FN)} \cdot 100\% \quad (4)$$

Keterangan :

1. TP (*True Positive*) menunjukkan jumlah data uji yang diklasifikasikan sistem ke dalam kategori x, dan semua data tersebut memang benar termasuk kategori x.

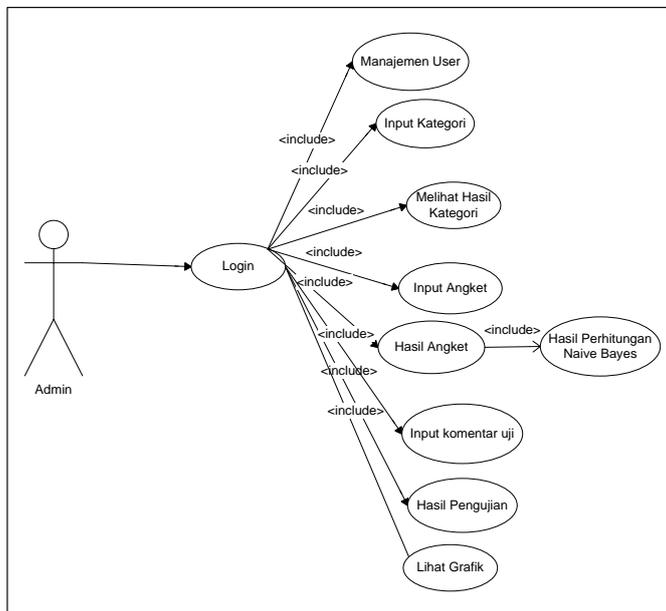
2. FP (*False Positive*) menunjukkan jumlah data uji yang tidak diklasifikasikan sistem ke dalam kategori x, tetapi seharusnya semua data tersebut termasuk kategori x.
3. FN (*False Negative*) menunjukkan jumlah data uji yang diklasifikasikan sistem ke dalam kategori x, tetapi seharusnya semua data tersebut bukan termasuk kategori x.
4. TN (*True Negative*) menunjukkan jumlah data uji yang tidak diklasifikasikan sistem ke dalam ketegori x, dan semua data tersebut memang bukan termasuk kategori x.

3. ANALISIS DAN PERANCANGAN

3.1 Perancangan UML

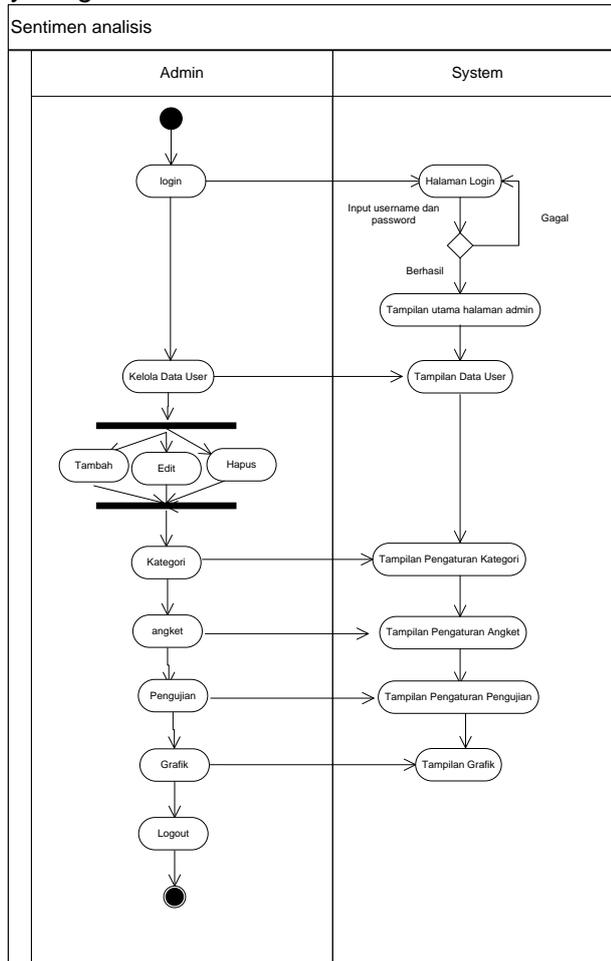
3.1.1 Diagram Use case

- Use case *Admin* :



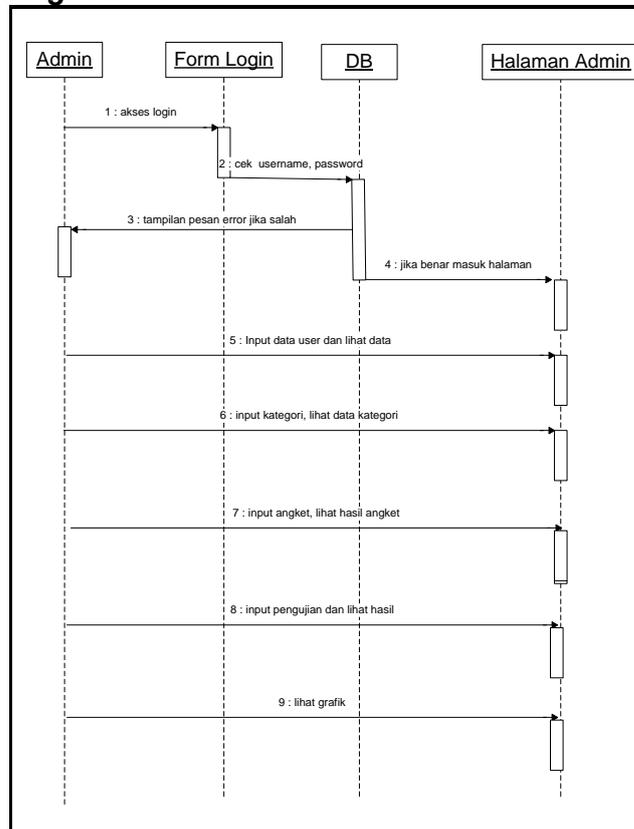
2.1.1 Activity Diagram

- Activity Diagram Untuk Admin :

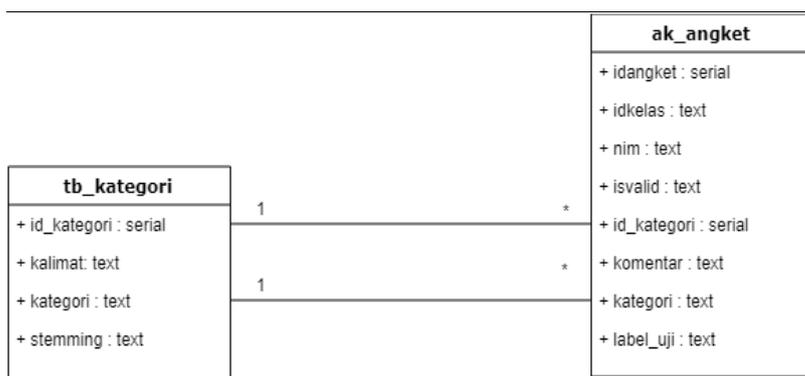


2.1.2 Sequence Diagram

- **Sequence Diagram Untuk Admin :**



2.2 Diagram Relasi Database

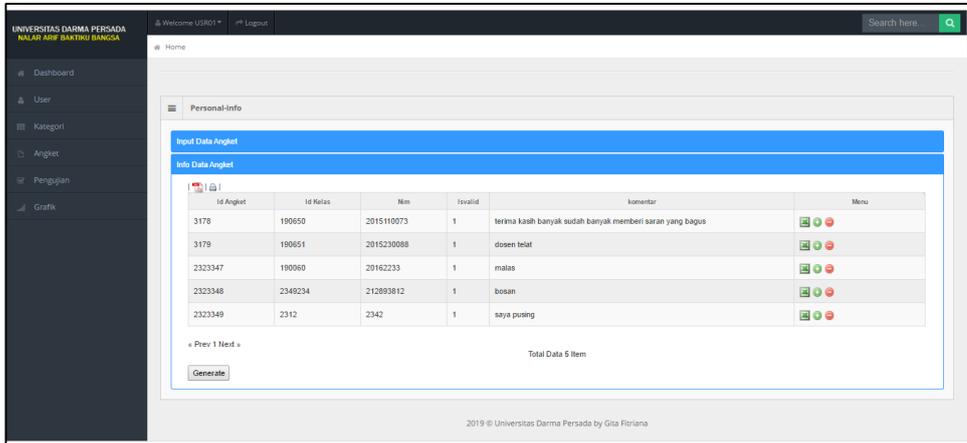


3. HASIL DAN PEMBAHASAN

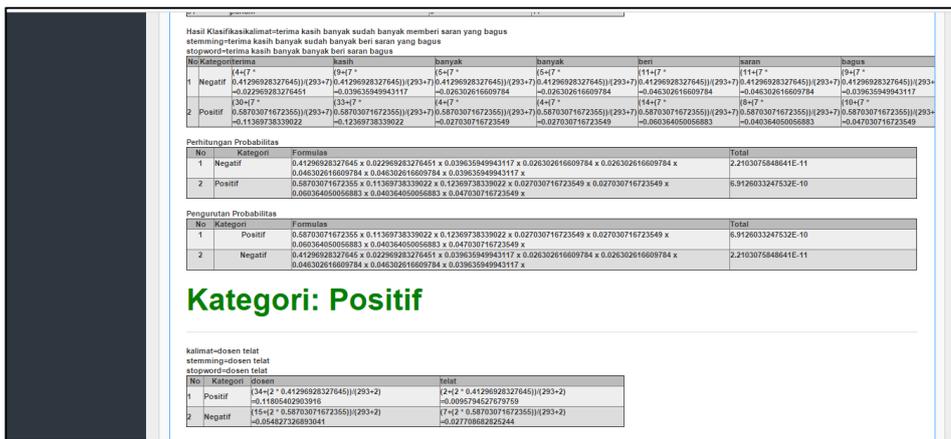
3.1 Desain Sistem

Interface Aplikasi web untuk admin

Halaman Info Data Angket



Halaman hasil generate data angket



Halaman Komentar Mahasiswa dan Nilai feedbacknya

No	Berbagai Komentar Mahasiswa	Sentimen penilaian
1	Tingkatkan kembali materi kuliah yang dapat memberikan motivasi untuk mahasiswa agar bisa menjadi manusia yang lebih baik	Positif
2	.	Positif
3	Pertahankan..	Positif
4	Baik	Positif
5	Sangat membimbing sekali dosen mata kuliah ini,	Positif
6	baik	Positif
7	Makasih pak	Positif
.....		
20	Terima kasih	Positif
21	good	Negatif
22	tidak ada	Negatif
23	Baik	Positif
24	bagus	Positif
25	Tidak Komentar	Negatif
26	Semoga lebih baik kedepan nya	Positif
27	Baik	Positif
28	baik	Positif
29	kinerja dosen sudah cukup baik, pertahankan	Positif

Nilai komentar : Positif 23 --> 3.97

3.2 Pembahasan

Contoh dianalisa feedback arah sentimen dari komentar berikut :

⇒ “Dosen sering telat”

Komentar di *stemming* berikut :

⇒ dosen sering telat

Komentar di *stopwords* berikut :

⇒ dosen sering telat

Selanjutnya komentar di pecah atau melakukan proses tokenisasi dari komentar melihat dari training yang ada sebelumnya, apakah data nya ada atau tidak berikut:

Tabel 4.2 tabel token

no	token	negatif	positif
1	dosen	38	38
2	sering	4	6
3	telat	7	2
token seluruh komentar training		920	1.393

total token seluruh komentar training :

$$\begin{aligned} \text{negatif} + \text{positif} &= 930 + 1.393 \\ &= 2.323 \end{aligned}$$

Seluruh komentar training :

- negatif = 146
- positif = 180
- total = 326

Menghitung prior sebagai berikut :

$$P(v_j) = \frac{\text{docs } j}{\text{training}}$$

$$P(\text{positif}) = \frac{180}{326} = 0,129217516$$

$$P(\text{negatif}) = \frac{146}{326} = 0,156989247$$

Menghitung probabilitas sebagai berikut :

$$P(a_i|v_j) = \frac{ni+1}{n+kosakata}$$

sehingga di dapat :

$$\begin{aligned} P(\text{dosen} | \text{positif}) &= \frac{38+1}{1393+2323} \\ &= \frac{39}{3716} = 0,010495156 \end{aligned}$$

$$\begin{aligned} P(\text{dosen} | \text{negatif}) &= \frac{38+1}{930+2323} \\ &= \frac{39}{3253} = 0,011988933 \end{aligned}$$

$$\begin{aligned} P(\text{sering} | \text{positif}) &= \frac{6+1}{1393+2323} \\ &= \frac{7}{3716} = 0,001883745 \end{aligned}$$

$$\begin{aligned} P(\text{sering} | \text{negatif}) &= \frac{4+1}{930+2323} \\ &= \frac{5}{3253} = 0,001537042 \end{aligned}$$

$$\begin{aligned} P(\text{telat} | \text{positif}) &= \frac{2+1}{1393+2323} \\ &= \frac{3}{3716} = 0,000807319 \end{aligned}$$

$$\begin{aligned} P(\text{telat} | \text{negatif}) &= \frac{7+1}{930+2323} \\ &= \frac{8}{3253} = 0,002459268 \end{aligned}$$

$$V_{\text{MAP}} = \operatorname{argmax}_{v_j \in \{a_1, a_2, a_3, \dots, a_n\}} P(v_j) :$$

sehingga di dapat :

$$V_{\text{MAP}} = \operatorname{argmax}_{\text{positif}} (0,010495156 \times 0,001883745 \times 0,000807319) \\ (0,129217516) = 2,06242281983 \text{ E}^{-9}$$

$$V_{\text{MAP}} = \operatorname{argmax}_{\text{negatif}} (0,011988933 \times 0,001537042 \times 0,002459268) \\ (0,156989247) = 7,11446149398 \text{ E}^{-9}$$

sehingga komentar tersebut kategori sentimen negatif karena jumlah probabilitas negatif lebih besar dari yang positif.

4. PENUTUP

4.1 Kesimpulan

Berdasarkan hasil pembahasan di atas dapat disimpulkan :

- 1 Sistem yang dibangun berhasil mengubah data dari bentuk kualifikasi (teks) diubah menjadi kuantifikasi (angka) secara otomatis.
- 2 Akurasi yang dihasilkan dari analisa ini dengan data sampel 326 komentar adalah sebesar 84%.
- 3 Kualitas keakuratan dan objektivitas model pada sistem ini ditentukan oleh kualitas dan kuantitas dataset yang digunakan. Semakin banyak dataset yang digunakan akan memberikan peluang akurasi yang lebih baik.

4.2 Saran

Beberapa saran untuk pengembangan lebih lanjut :

1. Penambahan jumlah data lebih agar menghasilkan hasil *training* dengan eror yang kecil serta tingkat akurasi prediksi mencapai nilai maksimal.

2. Selanjutnya perlu mengkombinasikan metoda yang digunakan pada penelitian ini dengan metoda lain yang berkaitan dengan text-mining dan analisa keakuratannya.

DAFTAR PUSTAKA

1. Jeffrey L Whitten, 2004, **Systems Analysis and Design Methods**, Pennsylvania State University, McGraw-Hill Irwin
2. Munnawar. 2005, **Pemodelan Visual dengan UML**, Graha Ilmu , Yogyakarta,
3. Feldman, R., & Sanger, J, 2007, **Text Mining Handbook, Advanced Approaches in Analyzing Unstructured Data**, New York, Cambridge University Press
4. Bunafit Nugroho, 2008, **Paduan Lengkap Menguasai SQL**, Jakarta , Mediakita,
5. A,S, Rosa dan Shalahuddin, M, 2011, **Rekayasa Perangkat Lunak (Terstruktur dan Berorientasi Objek)**, Bandung, Modula
6. A,M Hirin & Virgin, 2011, **Cepat Mahir Pemrograman Web dengan PHP dan MYSQL**, Pustakaraya , Jakarta
7. Saraswati, N,S, 2011, **Text Mining dengan Metode Naive Bayes Classifier dan Support Vector Machines untuk Sentiment Analysis**, Skripsi, Program Studi Teknologi Informasi Fakultas Teknik UGM Yogyakarta
8. F, Rozi, S, H, Pramono and E, A, Dahlan, "**Implementasi Opinion Mining (Analisis Sentimen) Untuk Ekstraksi Data Opini Publik pada Perguruan Tinggi**," Jurnal EECCIS, vol, 6, pp, 37-43, 2012
9. Liu, B, 2012, **Sentiment Analysis and Subjectivity, Synthesis Lectures on Human Language Technologies**, USA, Morgan & Claypool Publishers
10. Hermawati, F, A, 2013, **Data Mining**, Yogyakarta , Penerbit Andi
11. Rodiyansyah & Winarko, 2013, **Klasifikasi Posting Twitter Kemacetan Lalu Lintas Kota Bandung Menggunakan Naive Bayesian Classification**, *Indonesian Journal of Computing and Cybernetics Systems (IJCCS)*
12. M, Reza Faisal, 2014, **Seri Belajar Pemrograman , Pengenalan JQuery Untuk Pemula**, M, Reza Faisal , Banjarmasin
13. Jauhari Khairul Kawistara, Priyanto Hidayatullah, 2015, **Pemrograman Web**, Bandung, Penerbit Informatika
14. TIM EMS, 2015, **Kamus Komputer Lengkap**, Jakarta , Elex Media Komputindo,
15. Zaenal A, Rozi, 2015, **Bootstrap Design Framework**, Jakarta , Elex Media Komputindo
16. Hadna, N, M,, Santosa, P, I,, & Winarno, W, W, 2016, **Studi Literatur tentang perbandingan metode dengan proses analisis sentimen di twitter, Seminar Teknologi dan Komunikasi (SENTIKA)**
17. Mujihlawati, S, 2016, **Pre-processing Text Mining pada Twitter, Seminar Nasional Teknologi dan Komunikasi (SENTIKA)**
18. Mukhamad Masrur, 2016, **Pemrograman Web Dinamis Menggunakan Java Server Pages dengan Database Relasional MYSQL**, ANDI , Yogyakarta